**The Impact of Socioeconomic Factors on Urban Parking Meter Density**

By Jacky Lin, Kai Shang, Kexin Li, Yizhi Zhang

*Abstract: The paper analyzes the impact of socioeconomic factors (such as population, income and housing) on urban parking meter density. Through the analysis of the data, we use log function and the combination of variables (interaction). The results show that the heteroscedastic error and the low R² score, which indicate that the model does not fit the data well and there may be other significant predictors or non-Linear relationships.*

With the development of cities and the growth of population, parking problems have become the epitome of urban problems. In his insightful analysis of the country's parking situation, Henry Grabar shows how the need to store cars has made certain of our most pressing issues worse, such as the rising cost of housing and the impending global warming crisis (Grabar, 2023). Existing data indicate that the growth of urban population and frequent immigration flows in Canada have led to growing demand for housing, while there is a lack of affordable housing in both the rental and home purchase markets, and the risk of homelessness has been increasing (Teixeira, 2011). In Jones’ paper, he explores the logics of transit-oriented development in Vancouver and proves the rationality of high-density residential reconstruction, which illustrate how transit-oriented development regulations can lead to gentrification through the reconstruction of low-income rental housing after the establishment of a new rapid transit station (Jones, 2020).However, most of the research conducted by other scholars is from a macro perspective without discussing the impact of socioeconomic factors (such as population, income and housing) on specific measures of urban planning.

Urban infrastructure is a cornerstone of city planning, with parking meter distribution serving as a vital yet overlooked component. This study investigates how socioeconomic factors—median income, population, and dwellings—affect the allocation of parking meters. Based on the above, our research aims to answer the question: To what extent do population, income, and housing density affect parking meter allocation? By analyzing Vancouver's census and parking meter data, the results of this research will allow us to better understand the parking infrastructure distribution in different neighborhoods, which provides reference for the resource allocation of urban construction in the future. Meanwhile, it seeks to uncover inequities and offer insights into improving urban infrastructure planning for accessibility and fairness. Through the analysis of the data, we added the log function to stabilize the variance and reduce the impact of outliers. Meanwhile, considering the possibility of non-linearity relationship and the combination of variables that may produce compound or synergistic effects, we introduced the interaction term and performed an eyeball test on it. The heteroscedastic results and the low R² score indicate that the model does not fit the data well and there may be other significant predictors or non-Linear relationships.

The primary datasets utilized in this project is the Census Data. The census\_data\_parking data frame is the result of integrating parking meter data with the census data.

The integrated dataset, referred to as census\_data\_parking, was prepared and provided by the course instructor via a secured university drive.The dataset contains observations at the census-area level, with each row corresponding to a unique geographic unit. Key variables used in the analysis include:

**meters\_contained:** Total number of parking meters within a 5-meter radius of each census area.

**population:** Total population of the census area, representing neighborhood size and demand for infrastructure.

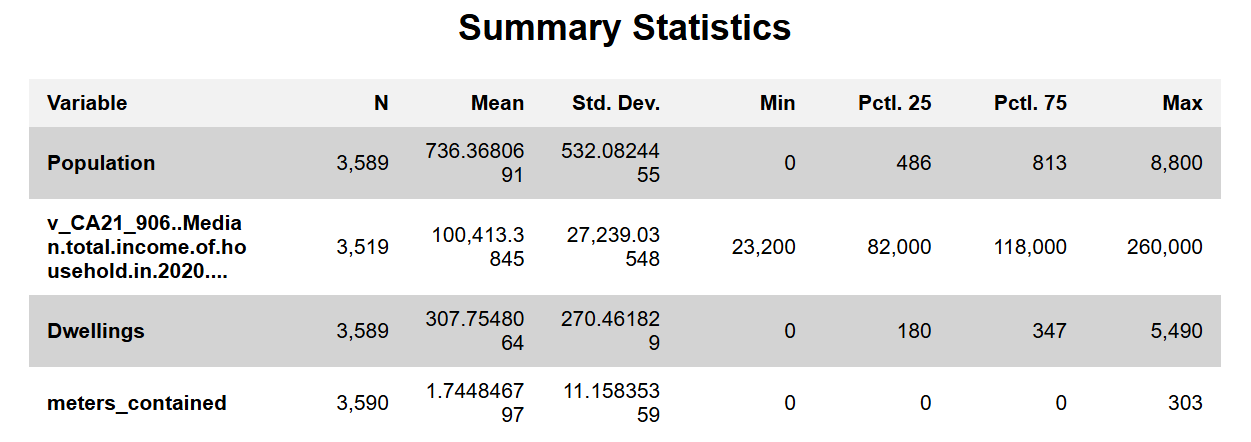
**income:** Median household income (2020), measuring the economic status of the area.

**dwelling**: Total number of residential dwellings in the census area, capturing housing density.

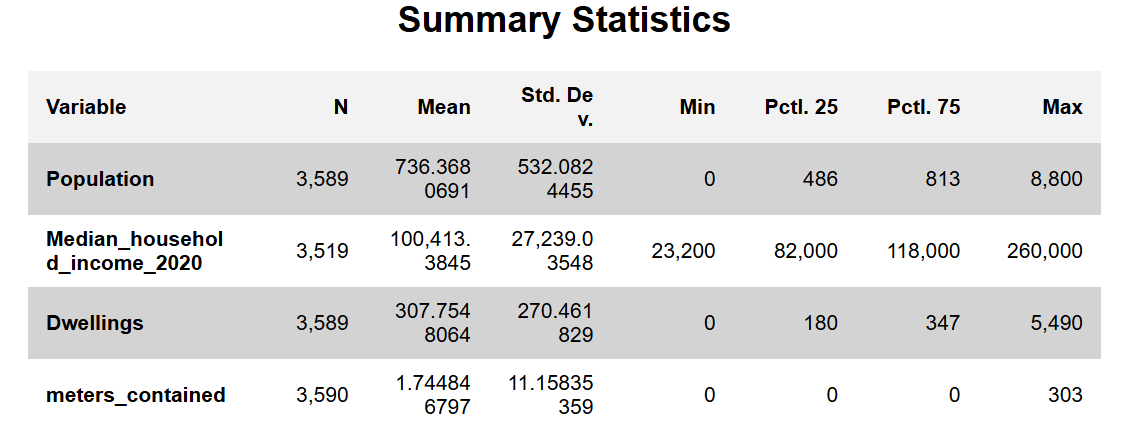
The main association being studied is the relationship between the density of parking meters in an area and socioeconomic and demographic parameters (such as population, income, and number of houses). This study attempts to determine whether wealthier or more populated areas are associated with a higher concentration of parking meters by modeling parking meter density as a function of population, income, and housing.

The dataset is structured as a tabular file. In the census\_data\_parking dataset, each row in the dataset represents a dissemination area in Vancouver, the columns record a variety of features such as housing characteristics, income information, demographic information that is important to include for the research. The dataset is organized for analysis, therefore no changes were necessary to the variables in this dataset.

The sample includes about 3,500+ of census locations in Vancouver. This sample size ensures a robust statistical power to detect meaningful relationships between parking infrastructure and socioeconomic variables. Vancouver is an urban city with a variety of neighborhoods that differ in infrastructure, density, and income. By analysing the census areas, we can provide an in-depth overview of the relationship between parking infrastructure and socioeconomic factors. These findings may possibly extend to different types of neighborhoods.



Renamed v\_CA21\_906..Median.total.income.of.household.in.2020.... to Median\_Household\_Income for clarity.



**Variable Descriptions**

**Independent Variables (X):**

**Population**

Description and justification: The total population residing in each census area. Population size reflects the demand for parking facilities. Areas with larger populations are likely to require more parking meters to accommodate residents and visitors, influencing the overall meter density.

Description and justification: Median household income for each census area, measured in Canadian dollars (CAD). Median income is a key indicator of the economic status of a neighborhood. Higher-income areas may have different parking needs and pricing strategies compared to lower-income areas, potentially affecting the number of parking meters placed.

Description and justification: Indicates number of residential units (dwellings) in each census area. The number of dwellings is indicative of residential density. Higher dwelling counts can lead to increased demand for parking, thereby affecting the number of parking meters needed in the area.

**Dependent Variable (Y):**

**meters\_contained**

Description and justification: This variable represents the total number of parking meters located within a 5-meter radius of each census area. meters\_contained serves as the primary measure of parking meter density in a neighborhood. Analyzing this variable allows us to understand how various socioeconomic factors influence the distribution and concentration of parking infrastructure across different areas.

This paper will use the multiple regression model to investigate the impact of socioeconomic factors on parking meter density in urban neighborhoods. mentioned above. It is expressed as:

Yi = β0 +β1X1i +β2X2i +β3X3i + ϵi​

The dependent variable (Yi), parking meter density (meters\_contained), represents the total number of parking meters within a 5-meter radius of each census area. The independent variables are population (X1i), median income (X2i), and number of dwellings (X3i).

|  |  |  |  |
| --- | --- | --- | --- |
| Variables | Model1 | Model 2 | Basic Model |
| Population | 0.002 (0.0003)\*\*\* | 0.001 (0.0004)\*\*\* | -0.014 (0.001)\*\*\* |
| Income | - | -0.00003 (0.00001)\*\*\* | 0.00003 (0.00001)\*\*\* |
| Dwellings | - | - | 0.034 (0.002)\*\*\* |
| Constant | 0.592 (0.317)\* | 3.857 (0.789)\*\*\* | -1.260 (0.817) |
| Observations | 3,589 | 3519 | 3,519 |
| R^2 | 0.006 | 0.012 | 0.086 |
| Adjusted R^2 | 0.005 | 0.011 | 0.085 |
| Residual Std. Error | 11.130 | 11.085 | 10.661 |
| F Statistic | 20.128 (1; 3587)\*\*\* | 21.127 (2; 3516)\*\*\* | 110.578 (3; 3515)\*\*\* |

When the result from model1 of the relationship between population (X1i) and dependent variable is checked, the problem is obvious. The value of R-squared is only 0.00558. This means that only 0.56% of the variation in parking meter allocation is explained by population size. Model1 is too poor to explain the relationship between population size and the allocation of parking meters. Then, the median income (X2i) is added to make the model better, because income is easy to connect to the vehicles. However, the outcome is still bad. Income not only has a weak relationship due to low R-squared(1.19%), but also has a negative relationship with parking meter allocation. It is a big shock. Although adding income improves the explanatory power slightly, the overall performance remains weak. The basic multiple linear regression model is setted. Compared to two X used before, dwelling density has a larger coefficient. It indicates a much better result as a strong positive relationship with parking meter allocation. However, dwelling do have a significantly better performance, the R-squared is still small. The model should be modified.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Coefficient | Std. Error | t-value | p-value |
| Intercept | -22.7158 | 3.4062 | -6.669 | < 0.001\*\*\* |
| log(Population) | -8.5667 | 0.3742 | -22.895 | < 0.001\*\*\* |
| log(Income) | 1.4360 | 0.2921 | 4.916 | < 0.001\*\*\* |
| log(Dwellings) | 8.8453 | 0.3449 | 25.647 | < 0.001\*\*\* |
| Metric | | Value | | |
| Residual Std. Error | | 3.852 | | |
| Multiple R^2 | | 0.2037 | | |
| Adjusted R2R^2R2 | | 0.203 | | |
| F-statistic | | 299.7 (df = 3, 3515) | | |
| p-value | | < 2.2e-16 | | |

There are many model specifications that can be used to ensure robustness and make the results be understood easily.

Initially, log-transformed specification is a good choice. It can be used to capture non-linear relationships, because it is especially effective where small changes in variables might have compounding effects on parking meter density. And it measures how proportional changes in population (X1i), income (X2i), or dwellings (X3i) affect parking meter density (Yi).

The adjusted R-square of the log-transformed model is ~0.203 as shown from the above table. Comparing the base model, the adjusted R-square is ~0.085. This significant improvement implies that the relationships between the predictors and rate are more accurately captured by the log transformation. Additionally, the residual standard error decreases substantially from ~10.66 to ~3.852, this further supports the significance of the log-transformed model. The use of log transformation helps stabilize variance and reduces the influence of outliers. It also captures the non-linear and proportional correlations between the predictors and parking meter density, which align better with how the data is distributed.

Second, this specification leads interaction terms into the initial model, such as the product of income and dwelling. This tests whether the combined effect of income and dwelling into a new variable influences parking meter density. As a method of creating a non-linear relationship of two Xs, it might potentially highlight areas where socioeconomic factors interact in unexpected ways.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Coefficient | Std. Error | t-value | p-value |
| Intercept | 6.895 | 1.151 | 5.989 | < 0.001\*\*\* |
| Population | -0.01900 | 0.001086 | -17.506 | < 0.001\*\*\* |
| Income | -0.00004942 | 0.00001096 | -4.510 | < 0.001\*\*\* |
| Dwelling | 0.01474 | 0.002732 | 5.395 | < 0.001\*\*\* |
| Income\*  Dwelling | 0.0000003209 | 0.0000003235 | 9.922 | < 0.001\*\*\* |
| Metric | | Value | | |
| Residual Std. Error | | 10.52 | | |
| Multiple R^2 | | 0.1101 | | |
| Adjusted R^2 | | 0.1111 | | |
| F-statistic | | 109.8 (df = 4, 3514) | | |
| p-value | | < 2.2e-16 | | |

By including the interaction term of income and dwelling, the adjusted R-square of the model improves from ~0.085 to ~0.1101. This improvement of R-square from the interaction term suggests a better model’s explanatory power. Furthermore, the residual standard error drops from around ~10.66 to about ~10.52. The slight reduction indicates a modest improvement in the model’s precision. Despite the drop being relatively small, the ANOVA test shows that the interaction terms still slightly increase the R-square, which suggest a slight improvement to the overall model.

Then the method to get a better model is coming. Combining the specification method used before, and having the optimal model.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Coefficient | Std. Error | t-value | p-value |
| Intercept | -23.11 | 11.58 | -1.995 | 0.04613 \* |
| log(Population) | -8.74 | 0.3848 | -22.974 | < 0.001\*\*\* |
| log(Income) | 1.679 | 1.087 | 1.544 | 0.12257 |
| log(Dwelling) | 8.832 | 0.3930 | 22.474 | < 0.001\*\*\* |
| Income | 0.002202 | 0.001038 | 2.202 | 0.22954 |
| Dwelling | -0.002674 | 0.001034 | -2.588 | 0.00970 \*\* |
| Income\* Dwelling | 3.332e-08 | 1.131e-08 | 2.945 | 0.00325 \*\* |
| Metric | | Value | | |
| Residual Std. Error | | 3.848 | | |
| Multiple R^2 | | 0.206 | | |
| Adjusted R^2 | | 0.2046 | | |
| F-statistic | | 151.8 (df = 6, 3512) | | |
| p-value | | < 2.2e-16 | | |

From the result, the combination of log-transformed variables and interaction term improves the model’s explanatory power, from the base model adjusted R-square of ~0.085 increases to ~0.2046. Compared to the log-transformed model, the adjusted r-square increased slightly from ~2.03 to ~0.2046, and the residual standard error decreased from ~3.852 to ~3.848. Unfortunately, the optimal model still has some problems. Dwelling, the most influential variable, has a negative coefficient here. This condition does not logically exist in the world. However, the model is too hard to modify now. Overall, this indicates a modest improvement in the model, making it the optimal solution for the current analysis of models.

The relatively low r-square (~0.2046) of the model suggests that there are possible other important predictors that our model did not include. As we are executing different models and analyses, several more serious issues arise that cause our models to fail.

The data set of population and dwelling introduces multicollinearity. Multicollinearity arises when independent variables in a model are correlated, making it difficult to interpret and will result in less reliable statistical inferences. In our model, the population and dwelling variables are highly correlated based on the analysis (in the optimal model’s result, the coefficients of population and dwelling are almost a pair of opposite numbers), this may be due to the higher populations usually leading to higher density of dwelling in urban development patterns.

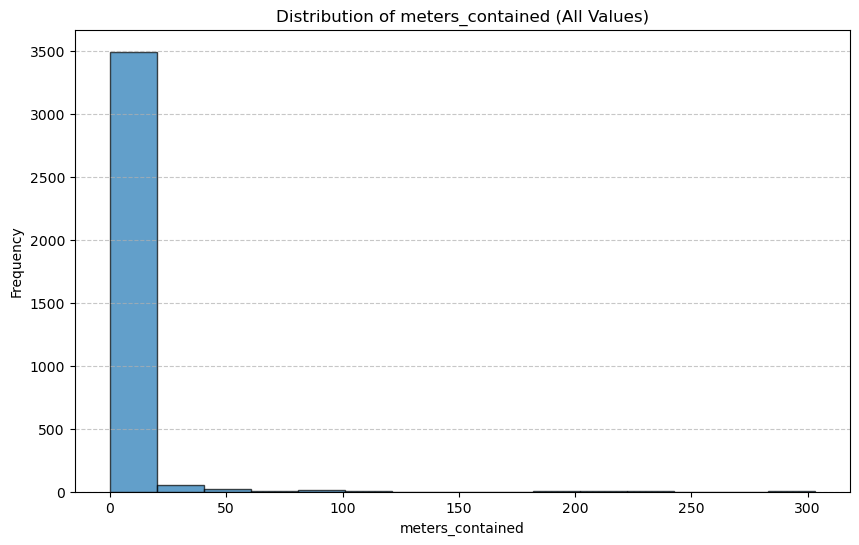
The presence of outliers was found in the dependent variable (Y). Outliers are data points that go far outside the mean value of a group, which will significantly affect the regression results. In this data set, the maximum is 303, 303 meters are contained in a specific 5-meter radius. It is absolutely not possible in real life. The large residuals are a result of outliers, this will raise the standard error. The outliers from meters-contained may be due to some areas with extremely high density of parking meters. Areas with high population and commercial activity will lead to some possible high outliers, and vice versa. Expanding further, not only the outliers, but also the Y’s dataset revealed the presence of heteroskedasticity in the sample. The presence of heteroskedasticity was found in our model when we performed the White test (eyeball test is also obvious). The unequal variance in our sample may cause our analysis results to be invalid.

Figure 1: Distribution of meters\_contained

Despite the low r-squared value indicates that the model is not comprehensive. However, the modified regression results show an increase in the value of the r-squared, indicating that variables such as population, housing, and income may remain key predictors of the control variables of the scale.

Additionally, the superbiased data variables are a huge problem in this paper. First, population, as a giant group of people, is a seriously biased variable. The paper wants to discuss the relationship between some variables and meters' destiny. The problem is what proportion of the whole population do have driver licenses. Does a person who does not know need to park a car? Also, the income variable is called Median income of the households. The problems come again, how many employees in these households? And what are the structures of these households (e.g. amount of kids)? Common sense tells us these datasets are fully biased.

Overall, the low r-squared value indicates several serious issues. Although the current model cannot fully explain parking meter density, it identifies certain meaningful and logically consistent relationships. Through this, we recognize the need for deeper modeling and data to properly depict the problem and complexity of urban development. Precisely picking the correct data for the model is the main key factor. Our future research should concentrate on choosing the right data and variables when building and enhancing our regression model.

As mentioned before, the estimated major issue of the model is multicollinearity between independent variables. Then, using a professional method, VIF, to prove that.

|  |  |
| --- | --- |
| Predictor | VIF Value |
| log(population) | 6.875012 |
| log(income) | 24.053800 |
| log(dwelling) | 10.263311 |
| income | 20.354389 |
| dwelling | 18.622568 |
| income:dwelling | 17.164123 |

According to the VIF result, most of the variables are indicating a high level of multicollinearity. And the main problem might come from the variable income. We estimate that there is a strong correlation between income and dwelling because income significantly influences the decision to buy a house. However, the income data is somewhat broad and should be broken down further. For example, it could be categorized by household income or income for individuals over 30. Luckily, The variable log(population) has relatively low VIF value, indicating that it is independent of other variables and contributes qualitate information to our model.

This research not only has practical significance and helps improve urban traffic and parking management, but also has high academic value. By exploring how socioeconomic factors such as population, income and housing density affect the distribution of parking facilities, it can provide a theoretical basis for transportation planning, urban management and policy formulation, promote sustainable urban development, and fill research gaps in related fields. But there are still some limitations. First, the research sample is limited to the Vancouver urban group, and future research can try to expand the sample scope to improve the generalizability of the results. Secondly, although linear regression was used in the research method, there is still room for improvement, such as introducing more variables or adopting a longitudinal research design to better explore causal relationships. Additionally, there are outliers in (meters\_contained), which show that the data plot is heteroscedastic with a low R² score. The result indicates that the model does not fit the data well and there may be other significant predictors or non- Linear relationships. Therefore, future research should further deepen the understanding of this issue through an interdisciplinary approach. It is hoped that this study can provide a useful reference for subsequent scholars and promote further development in this field.

**Citation：**

Grabar, H. (2023). *Paved Paradise*. Penguin.

Graves, J. (2024). Census Data Parking: Vancouver Dissemination Areas [Dataset: census\_data\_parking]. University of British Columbia. <https://ubcca-my.sharepoint.com/personal/jonathan_graves_ubc_ca/Documents/Teaching/Teaching%20Data/2024W1/output/parking/data_parking_nosf.rda>

Jones, C. E. (2020). Transit-Oriented Development and Suburban Gentrification: A “Natural Reality” of Refugee Displacement in Metro Vancouver. *Housing Policy Debate*, *33*(3), 1–20. <https://doi.org/10.1080/10511482.2020.1839935>

Teixeira, C. (2011). Finding a Home of Their Own: Immigrant Housing Experiences in Central Okanagan, British Columbia, and Policy Recommendations for Change. *Journal of International Migration and Integration / Revue de l’Integration et de La Migration Internationale*, *12*(2), 173–197. <https://doi.org/10.1007/s12134-011-0181-9>